

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-112576

(43)Date of publication of application : 23.04.1999

(51)Int.Cl. H04L 12/66  
 G06F 13/00  
 H04L 12/46  
 H04L 12/28  
 H04L 12/56

(21)Application number : 09-272832

(71)Applicant : HITACHI LTD  
 HITACHI INFORMATION TECHNOLOGY CO  
 LTD

(22)Date of filing : 06.10.1997

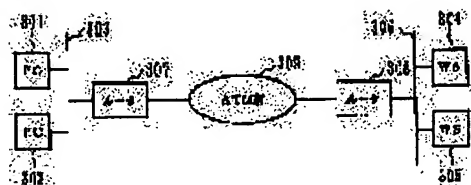
(72)Inventor : KIYONO TAKASHI  
 SHIN YOSHIFUMI  
 SAKOTA HIROYUKI

## (54) CONNECTION CONTROL METHOD FOR INTER-NETWORK SYSTEM

## (57)Abstract:

PROBLEM TO BE SOLVED: To prevent a channel utilization rate limited by flow control by the transmission control protocol TCP from being deteriorated in a network having a high transmission rate.

SOLUTION: Inter-network systems (routers) 307, 308 monitor TCP connection between adjacent hosts 301, 302, 303, 304, 305 and a host being a communication partner party of any of the adjacent hosts 301, 302, 303, 304, 305 and overcome a delay produced with the partner host by sending a reception acknowledgment segment to any of the adjacent hosts on behalf of the partner host when detecting deterioration in the channel utilization rate due to the TCP flow control. Furthermore, the routers 307, 308 store data sent from the adjacent hosts in their memories and conduct re-transmission acknowledgment with respect to a re-transmission request from the opposite host by using the data stored in the memories.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
 examiner's decision of rejection or application converted  
 registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
 rejection]

[Date of requesting appeal against examiner's decision of  
 rejection]

[Date of extinction of right]

BEST AVAILABLE COPY

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-112576

(43)公開日 平成11年(1999) 4月23日

(51)Int.Cl.<sup>6</sup>

識別記号

F I

H 0 4 L 12/66

H 0 4 L 11/20

B

G 0 6 F 13/00

3 5 1

G 0 6 F 13/00

3 5 1 A

H 0 4 L 12/46

H 0 4 L 11/00

3 1 0 C

12/28

11/20

1 0 2 E

12/56

審査請求 未請求 請求項の数3 O L (全 10 頁)

(21)出願番号

特願平9-272832

(22)出願日

平成9年(1997)10月6日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(71)出願人 000153454

株式会社日立インフォメーションテクノロジー

神奈川県秦野市堀山下1番地

(72)発明者 清野 崇

神奈川県海老名市下今泉810番地 株式会社日立製作所オフィスシステム事業部内

(74)代理人 弁理士 磯村 雅俊 (外1名)

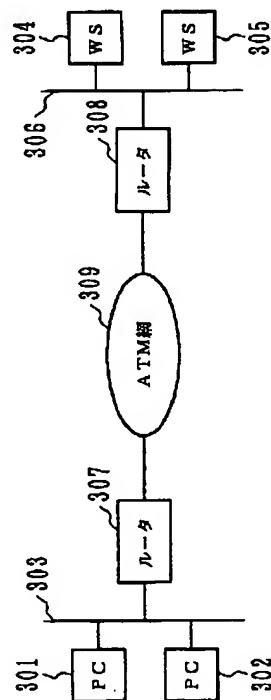
最終頁に続く

(54)【発明の名称】 インターネットワーク装置の接続制御方法

(57)【要約】

【課題】 高速なネットワークにおいて、TCPによるフロー制御により制限される回線使用率の低下を防ぐ。

【解決手段】 インターネットワーク装置（ルータ）307、308は、隣接ホスト301、302、304、305との通信相手となるホストとの間のTCPコネクションを監視し、TCPのフロー制御による回線使用率の低下を検出すると、相手ホストに代って隣接ホストに対して受信確認セグメントを送ることにより、相手ホストとの間に生じる遅延時間を克服する。また、ルータ307、308は、隣接ホストから送信されたデータをメモリ内に保存し、相手ホストの再送要求に対してメモリ内のデータを用いて再送応答する。



## 【特許請求の範囲】

【請求項1】 複数のネットワークを相互に接続し、TCP/IPを実装したネットワーク層レイヤ3のIPパケットの中継処理を行うインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、中継受信したデータパケットをIPの上位層のTCPレベルに渡し、

該TCPレベルで該インターネットワーク装置を経由するTCPコネクションの監視を行い、

経由するTCPコネクションを検出した場合、データ送信側ホストが自身と隣接しており、かつTCPのフロー制御による回線使用率が低下しているときには、データ受信側ホストに代って隣接する前記データ送信側ホストに対して受信確認セグメントを送ることを特徴とするインターネットワーク装置のコネクション制御方法。

【請求項2】 請求項1に記載のインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、TCPのフロー制御による回線使用率低下を検出するために、TCPコネクションに関連する回線の帯域幅、インターネットワーク装置自身で計算される回線使用率、RSTPによる帯域予約状況、及び隣接ホストとの間のラウンドトリップ遅延時間、ならびに隣接ホストと相手ホストとの間で合意されたウィンドウサイズから検出することを特徴とするインターネットワーク装置のコネクション制御方法。

【請求項3】 請求項1に記載のインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、隣接するホストからその相手ホストへ送信されるTCPセグメントの信頼性を保証するため、該隣接するホストに受信確認セグメントを送るとともに、該隣接するホストから送信されるデータセグメントをメモリに保存し、

該隣接ホストとの間の順序制御及びエラーチェックを行い、エラー時には該隣接ホストへの再送要求を行い、

前記相手ホストからの再送要求に対しては、保存したメモリ内のデータを用いて該再送要求に応答することを特徴とするインターネットワーク装置のコネクション制御方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、複数のネットワークを相互に接続して、TCP/IP (Transmission Control Protocol/Internet Protocol) を実装したインターネット装置において、高い回線使用率を実現するコネクション制御方法に関する。

## 【0002】

【従来の技術】 ある端末の最上位層のアプリケーションレイヤで生成された通信メッセージは、下位層のTCPレイヤおよびインターネットプロトコル（以下、IPと

記す）レイヤに伝達され、自端末と相手端末の各ドライバを介して送受信を行い、相手端末のIPレイヤから上位層のTCPレイヤに伝達され、さらにアプリケーションレイヤに伝達されることにより、アプリケーション相互間で通信が行われる。この場合、送受信両端末の間をメッセージが経由する他の端末では、ドライバからIPレイヤまでしかメッセージは伝達されることはない。従って、IPは信頼性のないデータグラム（IPデータ単位）の配送しか提供していない。この場合のホスト-ホスト間の信頼性は、IPの上位レイヤに位置するTCPが行う。TCPは、上位のアプリケーションレイヤに対してコネクションオリエンテッド（接続指向型）なサービスを提供し、通信の信頼性を保証する。TCPは、通信の信頼性確保のためにチェックサムによるパケットエラーのチェックやシーケンス番号を用いた正当な順序性の確保及び重複の排除を行うのみならず、データ転送の際には受信確認を行い、必要な場合には再送等の手段を用いて信頼性のあるサービスを提供するように努力する。TCPは、さらにフロー制御により相手が送信することができるデータ量のコントロールを行い、受信側が処理可能な範囲内の通信を行うと同時に、スロースタートアルゴリズムによるネットワーク内で起こる輻輳の回避を行う。

【0003】 TCPのフロー制御は、スライドする送信ウィンドウに基づいて行われる。このウィンドウ制御を用いることにより、送信側が確認をすることなく複数のデータスリムを連続して送信することが可能となり、その結果、スループットを向上させることができる。TCPは、本来、エンド-エンド間のプロトコルであって、送信側及び受信側の双方でのみ動作し、中継経路中のインターネットワーク装置等では動作しない。中継経路中のTCP制御機能に関する従来の技術としては、例えばパケットティア社のPacket Shaperがある。この装置は、送信側と受信側の間に入って、双方間のTCP制御パケットを奪い取り、その代りとなるTCP制御パケットを渡すことにより、帯域幅のコントロールを行い、パースト的に発生するトラフィックの分散を行うことにより、一定のサービスの品質（QoS: Quality of Service）を確保することが可能となる。また、例えば 特開平8-339354号公報に記載のネットワーク分散処理システムでは、システム間通信もシステム内通信と同じように、分散共有メモリ経由で行うことを基本とし、システム間で共通の分散共有メモリ空間を共有させる。送信プロセッサモジュールの分散メモリカップラが通信の方向を判断し、各プロトコルに合わせて送信側と受信側の共有メモリの同一アドレスロケーション間でコピーを行うことにより、WAN (Wide Area Network) を経由したシステム間通信のソフトウェアオーバーヘッドを削減し、処理効率を向上させることができる。

【0004】

【発明が解決しようとする課題】 前述の従来技術では、低速回線を用いた場合において、パースト的に発生するトラフィックの分散を主目的としているため、高速回線を用いたときに起こるTCPのフロー制御によるデータ送信効率の低減を抑制することはできない。TCPヘッダのウィンドウフィールドは、16ビット長しかないので、受信確認なしに一度に送ることができるデータのサイズは、約6万5千バイトに制限される。これは少なくとも約6万5千バイトにつき1回は受信確認のやり取りが行われ、その度毎に送信が中断することを意味している。この制限は、ネットワークの帯域幅は狭いときやラウンドトリップ遅延が小さいときには殆んど問題にはならないが、十分な帯域幅が確保された高速なネットワークの場合やラウンドトリップ遅延が大きな場合には、ウィンドウサイズがネックになるため帯域を十分に活用することができない。TCPの最近の実装においては、ウィンドウスケールオプションを用いてウィンドウサイズの倍率を指定することが可能になっているが、古いTCPの実装や安易なTCPの実装ではこのオプションを用いることはできない。また、ウィンドウスケールオプションを用いた場合においても、最大ウィンドウサイズは約1万6千バイトに制限されるため、将来出現するであろう超高速なネットワークにおいては、やはりウィンドウサイズがネックになりうる。このように、TCPはフロー制御を行うことにより、信頼性のあるエンド-エンドの通信を確保するが、フロー制御が制限になり、十分に回線の帯域を活用することができないという問題があった。そこで、本発明の目的は、このような従来の課題を解決し、ラウンドトリップ遅延の大きい高速なネットワークにおいても、ウィンドウサイズによる送信ネックを抑制し、回線の使用効率を向上させることが可能なインターネットワーク装置のコネクション制御方法を提供することにある。

【0005】

【課題を解決するための手段】 上記目的を達成するため、本発明によるインターネットワーク装置のコネクション制御方法では、複数のネットワークを相互に接続し、TCP/IPを実装したネットワーク層レイヤ3のIPパケットの中継処理を行うインターネットワーク装置において、インターネットワーク装置から直接配送されるネットワークに接続される隣接ホストと、その隣接ホストとの通信相手となる相手ホストとの間で行われるTCPコネクションを監視し、TCPのフロー制御による回線使用率の低下を検出するとともに、前記フロー制御により回線使用率が低下している場合には、相手ホストの代りに隣接ホストに対して受信確認セグメントを送ることにより、相手ホストとの間に生じる遅延時間を克服する。また、前記TCPのフロー制御による回線使用率の低下を検出する場合に、TCPコネクションに関連

する回線の帯域幅、インターネットワーク装置自身において計算される回線使用率、RSVPによる帯域予約状況、及び隣接ホストと相手ホストとの間のラウンドトリップ遅延時間、隣接ホストと相手ホストとの間で合意されたウィンドウサイズから、TCPのフロー制御による回線使用率低下を検出する。さらに、隣接するホストからその相手ホストへ送信されるTCPセグメントの信頼性を保証するため、隣接するホストから送信されるデータセグメントをメモリに保存し、隣接ホストとの間の順序制御及びエラーチェックを行い、エラー発生時には隣接ホストへの再送要求を行い、相手ホストからの再送要求に対しては、保存したメモリ内のデータを用いて再送要求に応答する。

【0006】

【発明の実施の形態】 以下、本発明の動作原理及びその実施例を、図面により詳細に説明する。

（動作原理） 本発明においては、送信ホストに隣接するインターネットワーク装置がTCPコネクションを監視し、各時刻における回線使用率が最適となるようにTCP受信確認パケットを相手受信ホストの代りに送出することにより、ウィンドウサイズの制限を克服する。通常、インターネットワーク装置の中継処理は、ネットワーク層レイヤ3のみで行われる。インターネットワーク装置は自身が保持しているレイヤ3（IPレベル）の中継処理テーブルの情報を基に受信したデータパケットをどのインターフェースから送出すべきかを判断し、該当するインターフェースに向けて送出する。本発明では、インターネットワーク装置は受信したデータパケットを一旦IPの上位層であるTCPレベルまで押し上げ、インターネットワーク装置を経由するTCPコネクションの監視を行う。TCPコネクションを検出できない場合には、再びIPレベルまでデータパケットが戻され、通常のIPレベルでの中継処理を行う。インターネットワーク装置を経由するコネクションを検出した場合、そのインターネットワーク装置は、上記コネクションに関して以下の条件が成立するか否かの検証を行う。

（条件1） データ送信側ホストは自身と隣接しているか否か、（つまり、自身から直接配送のネットワーク上に送信側ホストが接続されているか否か）

（条件2） TCPコネクションの関連する回線の帯域幅、帯域予約状況、ラウンドトリップ遅延時間、ウィンドウサイズからTCPの最大スループットを計算し、このTCPコネクションに関連した回線の持つ帯域を有効に使用することができないか否か、

【0007】 上記各条件のいずれかでも成立しない場合には、データパケットは再びIPレベルまで落されて、通常の中継処理を行う。また、上記各条件が両方とも成立した場合には、本発明におけるインターネットワーク装置は、このコネクション用のバッファを用意し、以後、このコネクションに関連するデータパケットをこの

バッファに保存するとともに、再び1レイヤまでデータパケットを戻し、通常の中継処理を行う。データパケットのバッファリングが開始された後に行われる送信側、受信側双方のTCP制御（データ転送以外のTCPセグメントの送受信）はインターネットワーク装置が肩代りして行う。つまり、送信側からのTCP制御はインターネットワーク装置がそのTCP制御に対する応答（さも、受信ホストからの応答であるかのように振舞う）を行い、また受信側からのTCP制御はインターネットワーク装置がその応答（さも、送信側からの応答であるかのように振舞う）を行う。また、必要な場合には、インターネットワーク装置側からTCP制御を送信または受信側に対してそれぞれの相手になり代って行う。上記処理を行うことにより、データセグメント送信後の受信確認待ち時間をラウンドトリップ時間からインターネットワーク装置と隣接する送信側ホスト間の2点間の送信時間に短縮することが可能になり、TCPのフロー制御による回線使用率低下を抑制することが可能になる。

【0008】データのバッファリングは、TCPの信頼性を保証するために必要である。例えば、インターネットワーク装置から受信側ホストに到達するまでの間に、送信データが何等かの理由によりデータ化けが起り、受信側ホストがチェックサムエラーを検出した場合を考える。このとき、受信側ホストのTCPは再送要求を送信側ホストに送出する。しかし、この時点で、このデータに対する受信確認セグメントは既にインターネットワーク装置から送信ホストに送られており、送信側ホストのTCPは送信ウィンドウを次へと移しているため、この受信側からの要求に対処することはできない。このような事態に備えて、このインターネットワーク装置ではTCPバッファリングを行い、保存してあるデータを用いて受信側からの再送に対処する。受信ホストからの再送要求に対して自身に保存してあるデータを基にして行うということは、バッファに保存されるデータの信頼性が保証されている必要がある。バッファ内のデータの信頼性を保証するために、このインターネットワーク装置は、バッファをベースとしたTCP処理（順序制御、エラー処理等）を行う。例えば、送信側から送られたTCPセグメントにおいてチェックサムエラーがあった場合には、それを受けたインターネットワーク装置は再送要求を送信ホストに要求を出し、正しいデータが保存されるように努力する。上記のことを実現するためには、インターネットワーク装置にTCPを実装する必要がある。本来、インターネットワーク装置はネットワークレイヤ3以下でのみ動作し、その上位層まで関知することはないため、IPの上位層であるTCPの実装は必須ではない。しかし、インターネットワーク装置の保守用にtelenetやftpのようなアプリケーションが必要になる場合が多く、殆どどのインターネットワーク装

置にはTCPが実装されている。そのため、本発明を実現するために既存のプロトコルスタックを改造するだけで、容易に実現できる。また、データの送信側及び受信側双方のプロトコルスタックの変更も必要なく、送信側に隣接するインターネットワーク装置のソフトウェアの変更のみで実現が可能である。

【0009】（実施例）図1は、本発明が適用されるルータ装置（インターネットワーク装置）のモジュール構成図である。本発明のルータ装置は、通信を中継する機能を有し、自身で送信および受信は行わない。ルータ装置は、パケットの中継処理やTCPの制御等のソフトウェア処理を行う主プロセッサ（CPU）101と、動作するソフトウェアをロードする目的の他に各種通信制御のバッファとして用いられる主記憶装置（メモリ）102と、このルータ装置上で動作させる制御ソフトウェアや構成定義情報を格納するためのディスク装置103と、複数の各種通信回線を制御するための回線制御機構105、106とから構成され、各装置は装置内バス104により接続されている。ここで構成定義情報は、ルーティングプロトコルに関する情報やアドレス情報等が登録されている。ルータ装置は、電源が投入されると、ディスク装置103から制御ソフトウェアが主記憶装置102にロードされ、このソフトウェアが主プロセッサ101により処理される。制御ソフトウェアは、ディスク装置103内に格納されている構成定義情報を読み込み、各種回線制御装置を動作させるとともに構成定義情報に従ったパケット中継処理を行う。

【0010】図2は、図1のルータ装置上で動作するソフトウェアの構成図である。通信の制御を行うソフトウェアは、図2に示す階層構造を形成している。最下層レイヤには、各種の異なる通信回線制御装置を制御するためのドライバ部201があり、その上位層には、IPデータグラムの中継処理、その他のIP処理を行うIP処理部202と、CSMA/CD（Carrier Sense Multiple Access with Collision Detection）方式のLAN（Local Area Network）において、IPアドレスとハードウェアアドレスとの間の対応付けを行うARP（Address Resolution Protocol）処理部203と、システム間の種々の調停のため制御コードをやり取りするためのICMP（Internet Control Message Protocol）処理部204と、実時間処理の必要な通信に対して帯域を予約するためのプロトコルであるRSVP（Resource Reservation Protocol）を処理するためのRSVP処理部205とが設けられる。さらにその上位層には、それぞれTCP（Transmission Control Protocol）、UDP（User Datagram Protocol）を処理するためのT

CP処理部206およびUDP処理部207と、このルータ装置間を経由するTCPコネクションを監視し、ウィンドウサイズによる通信ネックとなる場合に代替受信確認セグメントを送出して帯域幅の有効活用を行うコネクション制御部208とが設けられる。なお、TCPはコネクションを張って接続するのに対して、UDPはコネクションを張らずに制御セグメントを送出するのみである。さらに、最上位層には、他のルータ装置との経路情報のやり取りを行い、IP処理部202が参照する経路情報テーブルを更新するルーティング処理部209

と、アプリケーションプログラムであるftp(file transfer protocol)と、telnetとが設けられる。  
 【0011】図3は、本発明のルータ装置を用いたネットワーク構成例を示す図である。図3において、307及び308が、上記構成を持つ本発明のルータ装置である。2台のパーソナルコンピュータ(以下、PC)301、302及びルータ装置307は、100Mビット/秒の帯域幅を持つLAN303により接続され、また同様に2台のワークステーション(以下、WS)304、305とルータ装置308も100Mビット/秒のLAN306により接続される。2台のルータ装置307と308の間は、ATMスイッチ群により構成されるATMネットワーク網309に接続され、155Mビット/秒の帯域が確保されるが、ルータ装置間の距離は非常に離れており(例えば、東京-サンノゼ間)、ネットワークの遅延は非常に大きいものとする。PC302のTCPはウィンドウスケールオプションを実装しておらず、そのため最大ウィンドウサイズは65536バイトであるとする。いま、図3のPC301とWS304の間でビデオ会議アプリケーションを用いた通信を行っているものとする。このアプリケーションは、RSVPを用いて30Mバイト/秒の帯域予約を行っているものとする。このビデオ会議アプリケーションはエンド-エンド間の通信にUDPを用いているため、TCPのコネクションは存在しないものとする。この時、PC302とWS305の間において、TCPの上位アプリケーションであるftpを用いたファイル転送を行う場合を考える。

【0012】図4は、図3におけるWS305からPC302へデータ転送する際の通常の通信シーケンスチャートである。ここでは、通常のTCPを用いてデータ転送している場合を示している。まず、phase1では、WS305からPC302に対してシーケンス263000~328536のデータセグメントの送信を行い、phase2においてPC302からWS305に対してphase1で受信したデータの受信確認及び次に行われるphase3のウィンドウサイズ65536

バイトの指定を行っている。同様に、phase3では、次の65536バイトのセグメントの送信を行い、phase4でその確認を行っている。図4の右側に示したTt及びTdはそれぞれ65536バイトを送信するのに要した時間と、ネットワークのラウンドトリップ遅延時間を示している。このように、PC302とWS305の間には有効な帯域幅が70Mビット/秒存在する。すなわち、PC301とWS304の間では、LANの100Mビット/秒の帯域予約をRSVPを用いて行っているため、PC302とWS305の間の通信にはその帯域を差し引いた残りの帯域のみで行う必要があり、実際に使用できるのは100Mビット/秒から30Mビット/秒を引いた値となる。さらに、Tdの遅延が存在するため、70Mビット/秒のうちの(Tt/Td)×100%しかデータ転送には用いられない。この例のように、高速なネットワークで、かつ遅延時間が大きいときほど、ウィンドウサイズの制限がネットワークの回線使用率に重大な影響を及ぼす結果になる。

【0013】図5は、本発明の一実施例を示す代理受信確認を行う場合の通信シーケンスチャートである。前述のように、高速なネットワークで、かつ遅延時間が大きい場合には、ウィンドウサイズの制限がネットワークの回線使用率に大きな影響を及ぼす。この問題を克服するため、本発明においては、送信側に隣接するルータ装置308がTCPの受信確認セグメントをPC302の代りに発行し、WS305に送ることにより遅延時間による回線使用率の低下を抑制する。図5において、ルータ装置308はws305とPC302とのTCPコネクションを監視し、回線の帯域幅を効率的に使用するようにルータ装置308がPC302の代りにWS305に対して受信確認セグメントを発行すると同時に、PC302からWS305に対して送信されてくる受信確認セグメントを削除する。図5に示すように、ルータ装置308がPC302の代りに受信確認をWS305に送ることにより、ルータ装置308とPC302との間の送受信遅延を取り除くことが可能となるので、高いスループットを得ることができる。

【0014】本発明のルータ装置は、TCPのウィンドウサイズの制限により回線の持つ帯域幅を有効に使用することができない場合においてのみ代理で受信確認セグメントの発行を行い、その他の場合にはこの処理を行わない。ルータ装置308は、100Mビット/秒の帯域幅を持つLANインタフェースと155Mビット/秒の帯域幅を持つATMインタフェースの2つのインタフェースを装備しており、ここではそれぞれlan、atmと表現することにする。lan及びatmインタフェースの有効な帯域幅B1及びB2は、それぞれ次式

(1)、次式(2)で表わされる。

$$B1 = B_{atm} \times (R_a / 100) - B_r \quad \dots \dots \dots (1)$$

$$B2 = B_{lan} \times (R_l / 100) - B_r \quad \dots \dots \dots (2)$$

ここで、 $B_{atm}$ 及び $B_{lan}$ はそれぞれのインタフェースが持つ帯域幅、 $R_a$ 及び $R_l$ はそれぞれATM及びLANインタフェースの回線使用率、また $B_r$ はルータ装置において割り当てられたRSVPによる予約帯域を

$$B_b = \min(B_l, B_2) \dots \dots \dots (3)$$

【0015】一方、WS305とPC302の間のラウンドトリップ遅延時間を $T$ 、またWS305とPC302の間で合意したウィンドウサイズを $W$ とすると、このTCPコネクションでの最大スループット $B_w$ は次式

$$B_w = W/T \dots \dots \dots (4)$$

ここで、ウィンドウサイズ $W$ はウィンドウスケールオプションを考慮した値であっても差し支えない。もし、前式(4)の $B_w$ より前式(3)の $B_b$ の方が大きい場合には、TCPのウィンドウサイズによる回線使用制限が発生していることになる。一方、 $B_b$ の方が $B_w$ より大きい場合には、TCPのウィンドウサイズによる回線使用制限よりむしろ有効帯域幅がネックになっていると言える。そのためにルータ装置308では、WS305とPC302との間のTCPコネクションを監視し、ルータ装置が持っているその時点の回線使用率及び帯域予約状況から次式(5)を算出し、式(5)が真になる場合にのみ代理受信確認応答処理を行う。

$$(B_b > B_w) \dots \dots \dots (5)$$

これまでに述べた隣接するルータ装置による代理受信確認による方法は、WS305からPC302へのデータ送信で全てエラーがなく届いたときのみ有効に機能する。しかし、ルータ装置308からPC302へのデータ送信過程で新たにエラーが発生した場合には、回復する方法がない。すなわち、WS305は、ルータ装置308から受信確認セグメントを受信した後に、PC302からの再送要求が届いても、混乱してしまうと予想される。ここでWS305は、先に受信した受信確認セグメントをPC302から来たものであると思っているからである。このため、本発明においては、ルータ装置308がデータのバッファリングを行い、PC302からの再送要求に対応できるようにする必要がある。

【0016】図6は、本発明の他の実施例を示す再送要求に対処したシーケンスチャートである。図6では、WS305からシーケンス番号26300から32853までのTCPセグメントが送信され、ルータ装置308からPC302までの間にエラーがあったセグメントのみの再送要求を行っている。しかし、このデータに対する受信確認セグメントの送出は、ルータ装置308からWS305に対して既に行われているため、WS305に対して再送要求を送ることはTCPの動作に矛盾が起ってしまう。このような問題を解決するために、本発明においては、ルータ装置308はWS305から送られる全てのデータをルータ装置308上のメモリ102に保存し、PC302からの再送要求時にはこの保存さ

示す。WS305からPC302への通信時には、 $lan$ 、 $atm$ 双方のインタフェースを使用するため、この通信に有効な帯域幅 $B_b$ は次式(3)で表現できる。

れたデータを参照して再送要求に回答する。このデータは、PC302からのこのデータに対する受信確認セグメントが送られてくるまでメモリ102内に保存される。

10 【0017】一旦、PC302から再送要求が届くと、ルータ装置308は自身に貯えられたWS305からのデータを用いてそれまでに行われた通信の復旧を行い、復旧が終了するまでWS305には受信確認セグメントの送出は行わない。復旧処理が完了した時点で、ルータ装置308はWS305に対して受信確認セグメントの送出を行い、通信を再開させる。このように、ルータ装置308はWS305とPC302間のTCPコネクションにおいて行われるデータを自身のメモリ内に保存することにより、再送要求に対応させている。しかし、これはルータ装置308内のメモリに保存されるデータにエラーがないこと、そして順序が正しいことを前提としている。このことを保証するために、ルータ装置308は通常のTCPによるデータ転送と同様のことをWS305との間で行い、保存するデータの妥当性を保証する。例えば、WS305からPC302宛に送信されたデータセグメントにエラーを見つけた場合には、ルータ装置308はそのままPC302に対してデータを渡さず、WS305に対して再送要求を行い、エラーの修復を行う。また、順序通りに届かなかったデータセグメントに対しては、自身のメモリ内においてデータ順序の再構成を行う。

【0018】図6では、WS305からルータ装置308にはデータ1～5を正常に転送したので、ルータ装置308はメモリにデータ1～5をバッファリングするとともに、PC302に対して転送するとともに、ルータ装置308からWS305に対して代理受信確認を行う。しかし、ルータ装置308とPC302との通信経路中にエラーが発生したため、PC302からデータ4について再送要求が返送された。その時点では、WS305からデータ6～10の送信が終了し、ルータ装置308から代理受信確認が返送されている。そして、WS305からはデータ11～15がルータ装置308に対して転送されている。ルータ装置308は、このデータ11～15をメモリに格納する。ルータ装置308は、その時点で以降の代理受信確認を中止し、PC302に対してメモリに保存していたデータ4をPC302に再送する。PC302からAck6が返送されると、ルータ装置308は次のデータ6～10までをメモリから取り出して送信する。PC302からAck11が返送されると、ルータ装置308はメモリに保存しているデー



タ11~15をPC302に転送する。PC302からAck16が返送されると、ルータ装置308はメモリにデータが残っていないため、その時点でWS305に対して代理受信確認を行う(Ack16)。

【0019】図7は、本発明の一実施例を示すルータ装置のバケット中継処理のフローチャートである。このフローは、図1に示したルータ装置のLAN回線制御機構105及びATM回線制御機構106からCPU101へのハードウェア割り込み処理により起動される(START)。図3におけるLAN306からのデータ受信により、図2に示すルータ装置308のLANDライバ部201がLANフレームの受信処理を行い(ステップ701)、そのデータをIP処理部202に渡す。IPデータグラムを受けたIP処理部202はIPの受信処理を行い(ステップ702)、中継処理を行う前に上位レイヤであるTCP制御部206に引き渡す。このTCP処理部206が本発明におけるTCPコネクション制御を行う部分である。TCP制御部206はコネクションの監視を行い、TCPコネクション制御が有効な場合においてのみバッファリング、TCPの肩代り処理を行う(ステップ703)。バッファリングや肩代り処理を行った後、IP処理部202に渡し、IP中継処理を行い(ステップ704)、さらにLANDライバ部201に引き渡して、送信処理を行う(ステップ705)。一方、TCPコネクション制御が必要ない場合には、再びIP処理部202にIPデータグラムが戻され、ルーティングテーブルの情報をもとにIP中継処理を行い(ステップ704)、適切なインタフェースからIPデータグラムが送出される(ステップ705)。

【0020】図8は、図7におけるTCP制御部の詳細フローチャートである。TCP制御部206では、まずTCPのデータであるか否かを判断し(ステップ801)、隣接ホストからの送信であるか否かのチェックを行う(ステップ802)。どちらかが該当しない場合には、TCPコネクション制御を行わずにそのままIPデータグラムの中継処理に戻す(ステップ811)。どちらにも該当する場合には、既にコネクション制御を行っているか否かを判別し(ステップ803)、コネクション開始前のものであれば、コネクション確立用のTCPセグメントであるか否かの検査を行う(ステップ804)。その結果、コネクション確立用のTCP制御セグメントである場合には、前式(5)つまり(Bb>Bw)の判定を行い(ステップ805)、真の場合にはコネクションの登録及び監視を開始する(ステップ806)。一方、既にコネクションの登録が行われていた場

合には(ステップ803)、データのバッファリングを行った後(ステップ807)、このバッファをベースとした代替TCP処理を行う(ステップ808)。このバッファベースのTCP処理では、これまでに述べた代理TCP受信セグメントの送出やバッファ内の順序制御、再送要求、再送要求に対する代理応答等が行われる。TCP処理が終了した後、必要なデータをIP送信処理に渡し(ステップ809)、不要になったバッファ内のデータの削除を行い、バッファを開放して処理を終了する(ステップ810)。

【0021】

【発明の効果】以上説明したように、本発明によれば、送受信を行うホストのプロトコルスタックを変更することなく、高速なネットワークにおいてTCPによるフロー制御により制限される回線使用率の低下を防ぐことができ、高いスループットを得ることができる

【図面の簡単な説明】

【図1】本発明が適用されるルータ装置のハードウェア構成図である。

【図2】図1におけるルータ装置のソフトウェア構成図である。

【図3】本発明が適用されるネットワーク構成例を示す図である。

【図4】本発明を使用しない通常のTCP動作のシーケンスチャートである。

【図5】本発明の一実施例を示す代理受信確認動作のシーケンスチャートである。

【図6】本発明の他の実施例を示す再送発生時の動作シーケンスチャートである。

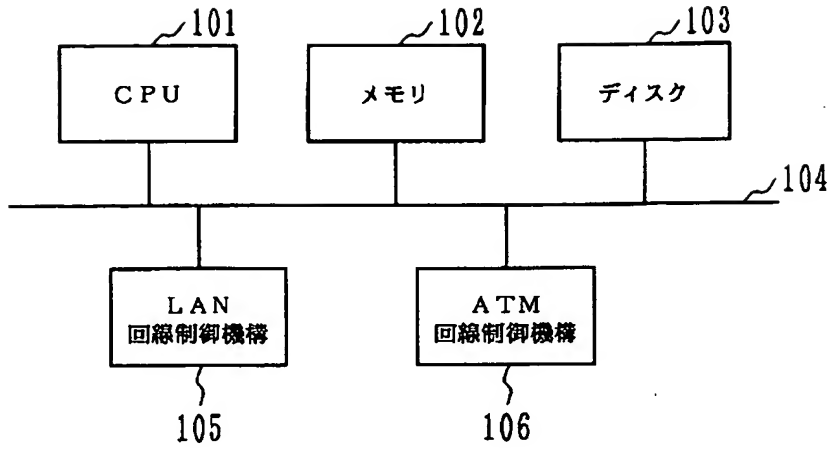
【図7】本発明のルータ装置における中継処理のフローチャートである。

【図8】図7におけるTCP制御部の処理フローチャートである。

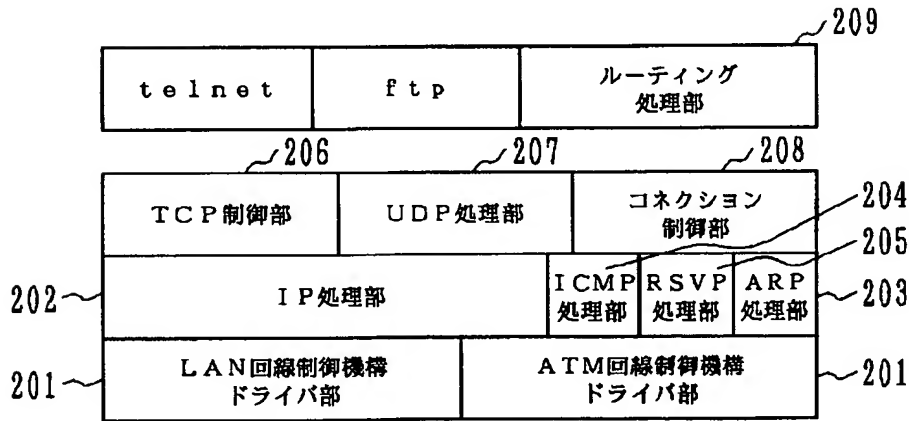
【符号の説明】

101…CPU、102…メモリ、103…ディスク、104…内部バス、105…LAN回線制御機構、106…ATM回線制御機構、201…LAN回線制御機構ドライバ部、ATM回線制御機構ドライバ部、202…IP処理部、203…ARP処理部、204…ICMP処理部、205…RSVP処理部、206…TCP処理部、207…UDP処理部、208…コネクション制御部、209…ルーティング処理部、301、302…PC、304、305…WS、307、308…ルータ装置、309…ATM網、303、306…LAN。

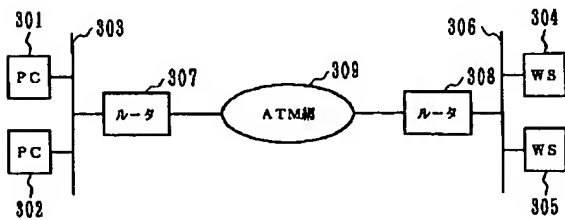
【図1】



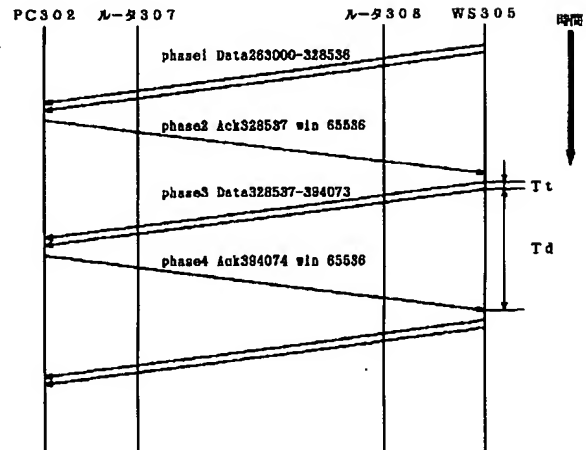
【図2】



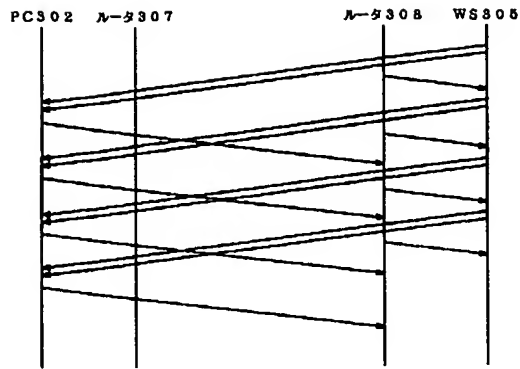
【図3】



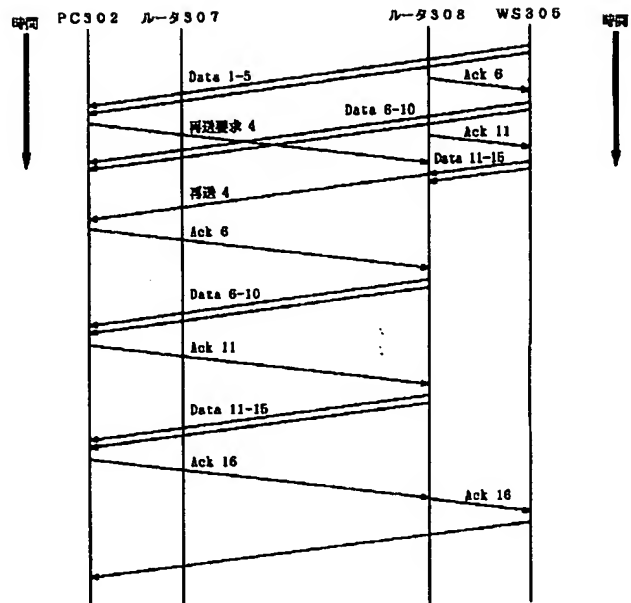
【図4】



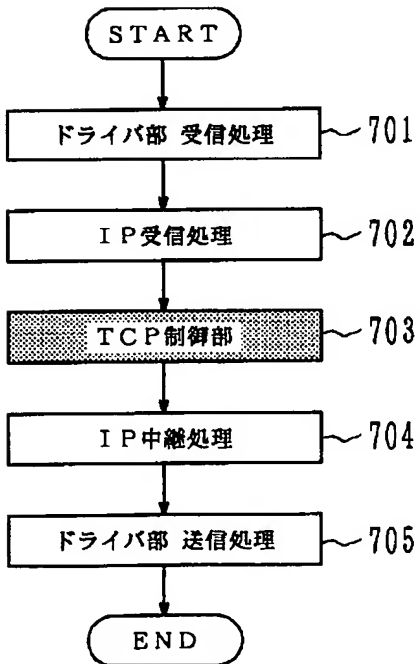
【図5】



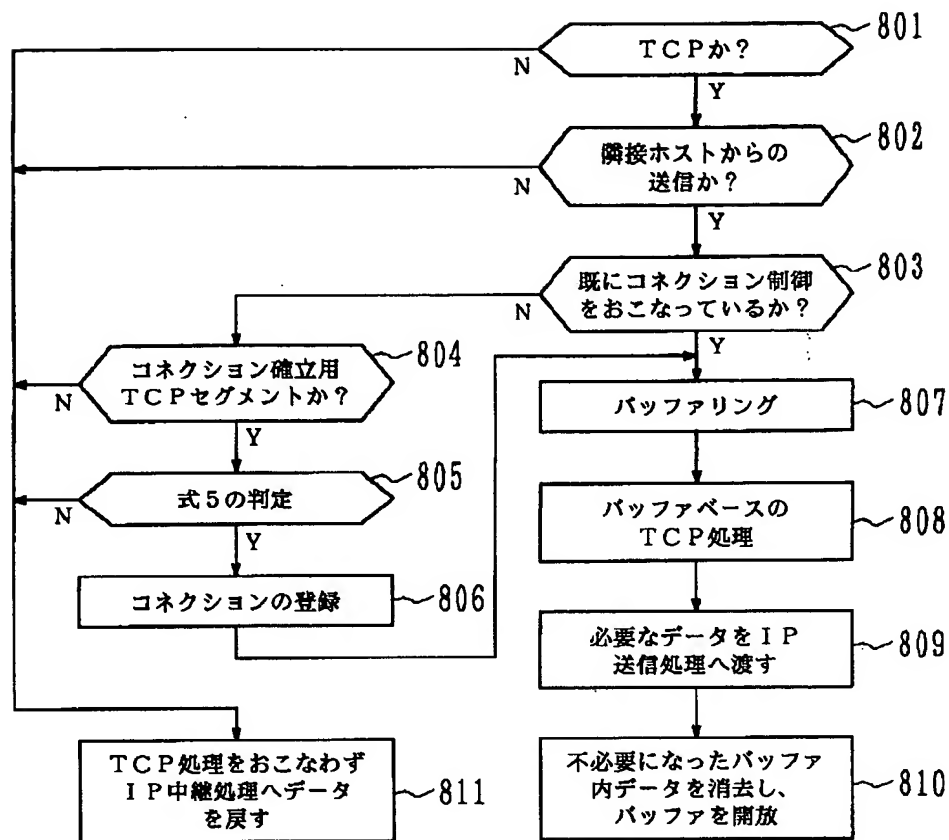
【図6】



【図7】



【図8】



フロントページの続き

(72)発明者 新 善文  
神奈川県海老名市下今泉810番地 株式会  
社日立製作所オフィスシステム事業部内

(72)発明者 迫田 博幸  
神奈川県秦野市堀山下1番地 株式会社日  
立インフォメーションテクノロジー内

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**